ベイズ推定を用いた画像パターンからの支配方程式の選択

義永　那津人

WPI-AIMR, 東北大学
MathAM-OIL, 産総研

共同研究者：
徳田　悟（九州大学）

National Institute of Advanced Industrial Science and Technology
AIST

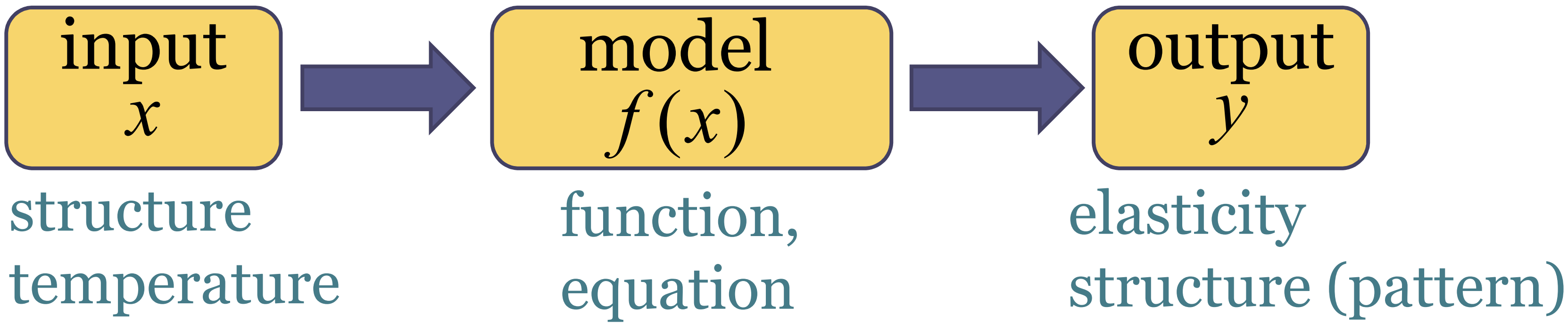
Mathematics Advanced Materials -OIL

AIMR

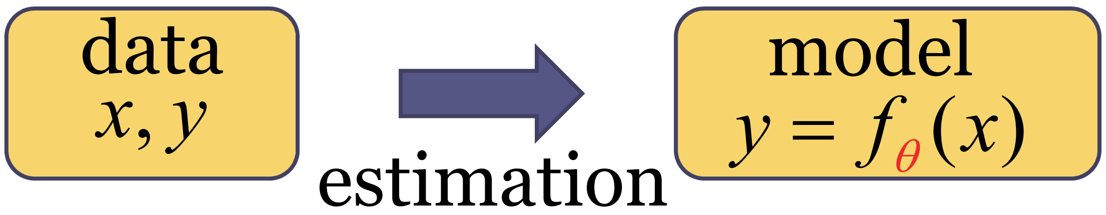
TOHOKU UNIVERSITY

Machine Learning, Data-Driven Science, Inverse Problem, Optimisation

- forward problem



- inverse problem



We do not know the true model...

- Parametrised model
  - polynomials $f_\theta(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \cdots$
  - Fourier expansion $f_\theta(x) = \theta_1 \cos x + \theta_2 \cos 2x + \cdots$
  - deep learning $f_{w,b}(x) = \sigma\left(W_n \sigma\left(\ldots\sigma(W_1 x + b_1)\right) + b_n\right)$

- Learning
  - supervised $\{x, y\} \to y = f_\theta(x)$
  - unsupervised $x \to f_\theta(x)$ such that x is linearly discriminable
  - generative $\mathcal{N}(z) \to p_\theta(x)$
    - normal dist. to Boltzmann dist.
    - deep learning

interpretability or predictability

Contents

- Discovery of a governing equation from time series data (Nathan Kutz)
- Inverse structural design for molecular dynamics and Monte Carlo simulations (Thomas Truskett)
- Bayesian modelling of partial differential equations from one snapshot of pattern (arxiv:2006.06125)

Discovery of a Model (Brunton-Proctor-Kutz (2016))

- Ordinary differential equations

$$\dot{x}(t) = f(x(t)) \Rightarrow \dot{x}(t) = f_\theta(x(t))$$

$$\dot{x}(t) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \cdots$$

- Partial differential equations

$$\partial_t u(x,t) = f(u(x,t)) \Rightarrow \partial_t u = f_\theta(u, \partial_x u, \cdots)$$

$$\partial_t u = \theta_0 + \theta_{1,0} u + \theta_{2,0} u^2 + \theta_{1,1} \partial_x u + \theta_{1,2} \partial_{xx} u \cdots$$

Can we estimate a model from data?

⇒ Parametrise r.h.s. by θ and perform regression

Sparse Identification of Dynamical systems (SINDy, PDE-FIND)

model

$$\partial_t u = \theta_0 + \theta_{1,0} u + \theta_{2,0} u^2 + \theta_{1,1} \partial_x u + \theta_{1,2} \partial_{xx} u \cdots$$

data

$$\{x_i, t_i, u_i(x_i, t_i)\}_{i=1,\dots N}$$

$$\partial_t \mathbf{U} = \Xi(\mathbf{U}) \cdot \Theta$$

candidate terms

time derivatives

$$\begin{pmatrix} \partial_t u_1 \\ \partial_t u_2 \\ \vdots \\ \partial_t u_N \end{pmatrix} = \begin{pmatrix} u_1 & \partial_x u_1 & u_1^2 & \cdots \\ u_2 & \partial_x u_2 & u_2^2 & \cdots \\ u_3 & \partial_x u_3 & u_3^2 & \cdots \\ \vdots & \vdots & & \vdots \end{pmatrix} \cdot \begin{pmatrix} \theta_{1,0} \\ \theta_{1,1} \\ \theta_{2,0} \\ \vdots \end{pmatrix}$$

parameters

sparse regression

$$\Theta = \arg\min_{\Theta} \left[\left| \partial_t \mathbf{U} - \Theta \cdot \Xi(\mathbf{U}) \right|^2 + \lambda \left| \Theta \right| \right]$$

sparsity

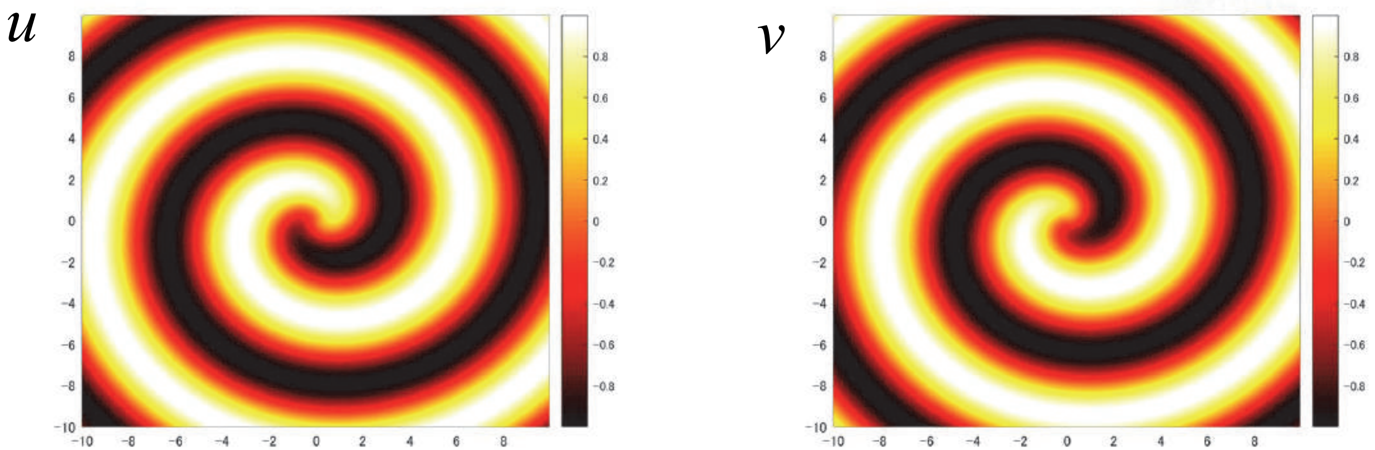
Example: reaction-diffusion equations

ground-truth model

Rudy-Brunton-Proctor-Kutz (2017)

$$\partial_t u(x,t) = u - (u^2 + v^2)u + (u^2 + v^2)v + 0.1\Delta u$$

$$\partial_t v(x,t) = v - (u^2 + v^2)v - (u^2 + v^2)u + 0.1\Delta v$$



estimation

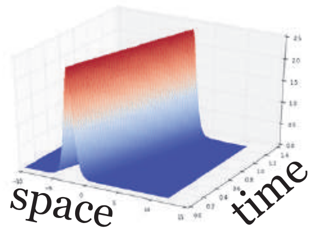
$$\partial_t u(x,t) = 1.00u - 1.00u^3 + 1.00v^3 - 1.00uv^2 + 1.00u^2 v + 0.100\partial_{xx} u + 0.100\partial_{yy} u$$

$$\partial_t v(x,t) = 1.00v - 1.00v^3 + 1.00u^3 - 1.00uv^2 - 1.00u^2 v + 0.100\partial_{xx} v + 0.100\partial_{yy} v$$

sparsity

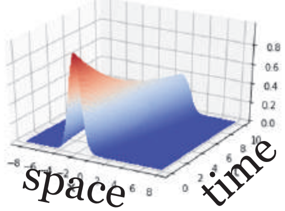
$$+0 \left\{ \begin{matrix} u^2 \\ uv \\ \partial_x u \\ \vdots \end{matrix} \right\}$$

Discovering a Model



Korteweg–de Vries (KdV) equation

$$\partial_t u + 6u\partial_x u + \partial_{xxx} u = 0$$

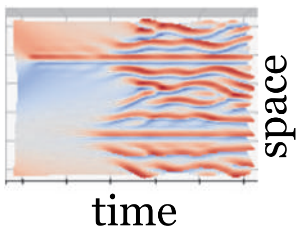


Burgers equation

$$\partial_t u + u\partial_x u = \epsilon \partial_{xx} u$$

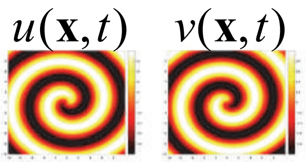
Nonlinear Schrödinger

$$i\partial_t \psi + \frac{1}{2}\partial_{xx}\psi + \psi|\psi|^2 = 0$$

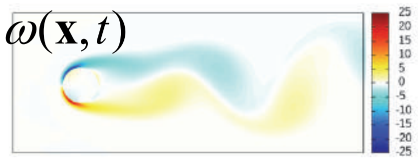


Kuramoto–Sivashinsky equation

$$\partial_t u + \frac{1}{2}(\partial_x u)^2 + \partial_{xx} u + \partial_{xxxx} u = 0$$



reaction-diffusion equation



Navier-Stokes equation

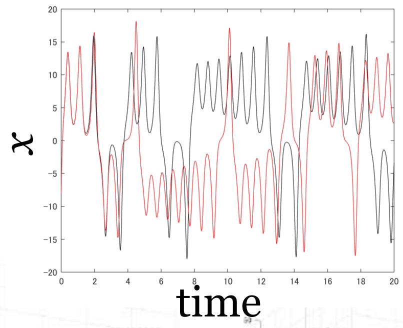
$$\mathrm{Re}\left(\partial_t \omega + (\mathbf{v}\cdot\nabla)\omega\right) = \Delta\omega$$

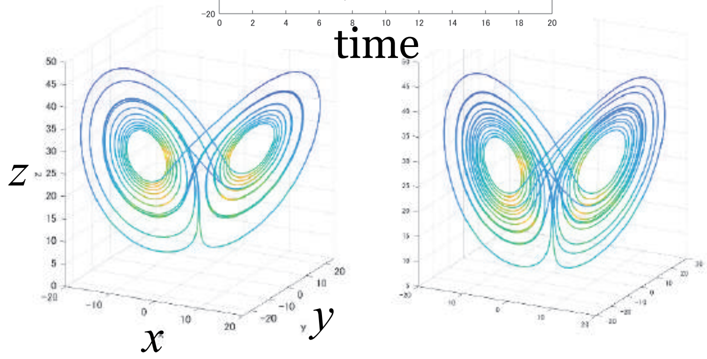
Lorentz model

$$\dot{x} = \sigma(y - x)$$

$$\dot{y} = x(\rho - z) - y$$

$$\dot{z} = xy - \beta z$$





Brunton-Proctor-Kutz (2016)

Rudy-Brunton-Proctor-Kutz (2017)

Remarks

- Equation discovery for time series has a long history
  - system identification (Crutchfield-McNamara 1987, Bär et al. 1999, Muller-Timmer 2004)
  - Finding an equation of motion for a double-pendulum from data (Bongard-Lipson 2007, Schmidt-Lipson 2009)
  - Sparse regression is simple and reasonably powerful, and looks widely applicable.

- Application
  - time-dependent parameters (group sparsity)
  - hidden variables

- Not always work
  - Precise evaluation of time and spatial derivatives is necessary.
  - Not robust against noise.
  - Need a lot of data (enough information including relaxation process).
  - What you get is a normal form, not necessarily a ground-truth equation.
  - Assume the state noise, not observation noise

Inverse Structural Design

- forward problem
  - molecular dynamics simulations, Brownian dynamics, Monte Carlo

initial condition $x_i(0)$ → model $x_i = -\nabla U(\{x_i\}) + \xi_i$ → stationary state $p(\{x_i\})$

Brownian dynamics
+ potential energy

- inverse problem

target structure $x_i \sim p_{\mathrm{tgt}}(\{x_i\})$ → (estimation) → model $U_\theta(\{x_i\})$

$$p_{\mathrm{tgt}}(\{x_i\}) \approx p(\{x_i\} \mid \theta)$$

Inverse Structural Design

Boltzmann distribution

$$p\left(\{x_i\}\mid\theta\right)=\frac{e^{-\beta U_\theta(\{x_i\})}}{\mathcal{Z}}$$

partition function (normalisation)

$$\mathcal{Z}=\prod_i\int e^{-\beta U_\theta(\{x_i\})}dx_i$$

Kullback-Leibler (KL) divergence

$$D_{KL}\left(p_{\text{tgt}}(\{x_i\})\,\|\,p(\{x_i\}\mid\theta)\right)=\int p_{\text{tgt}}(\{x_i\})\ln\frac{p_{\text{tgt}}(\{x_i\})}{p(\{x_i\}\mid\theta)}dx_i$$

Find θ satisfying $\partial_\theta D_{KL}\left(p_{\text{tgt}}(\{x_i\})\,\|\,p(\{x_i\}\mid\theta)\right)=0$

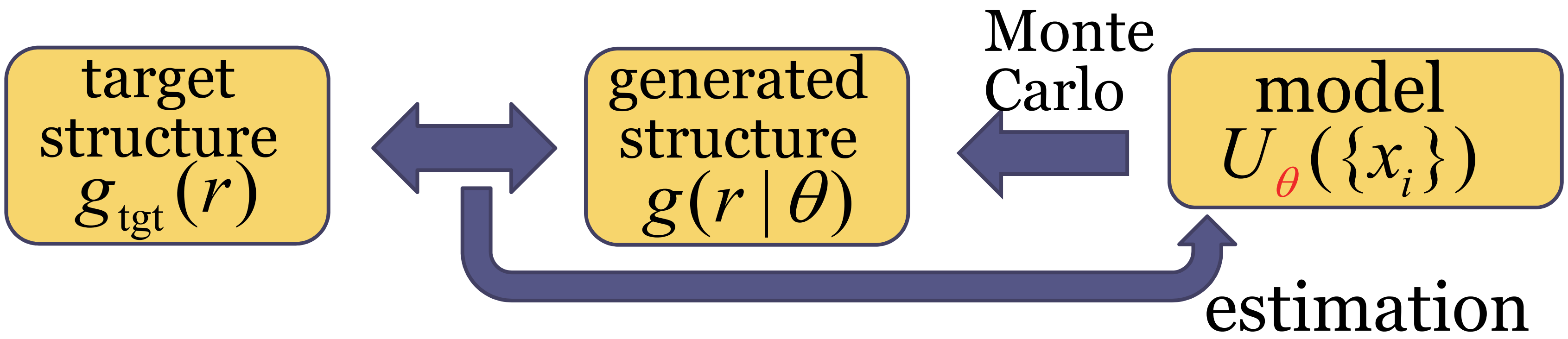
$$\theta^{(k+1)}=\theta^{(k)}-\alpha\partial_\theta\left[\left\langle\ln p(\{x_i\}\mid\theta)\right\rangle_{P_{\text{tgt}}(\{x_i\})}\right]_{\theta=\theta^{(k)}}$$

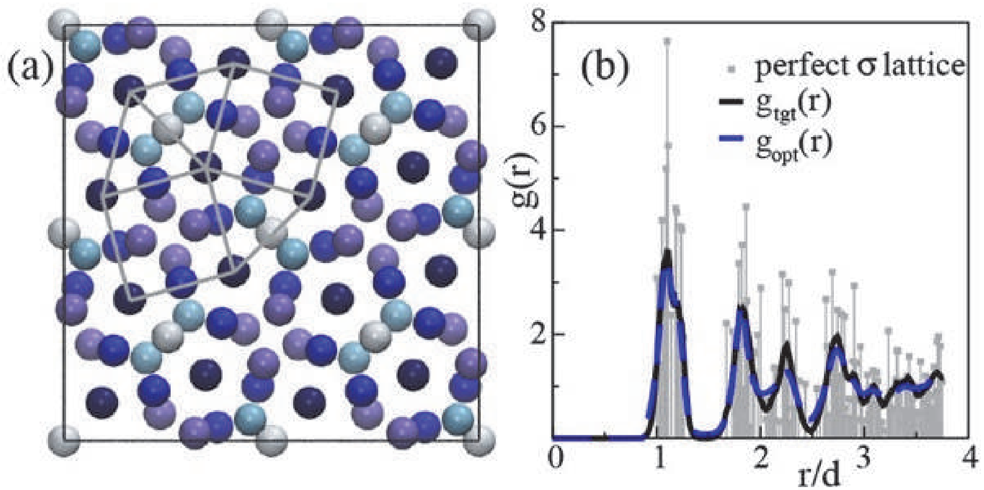
$$\left\langle\partial_\theta\beta U(\{x_i\}\mid\theta)\right\rangle_{P(\{x_i\}\mid\theta)}-\left\langle\partial_\theta\beta U(\{x_i\}\mid\theta)\right\rangle_{P_{\text{tgt}}(\{x_i\})}$$

Inverse Structural Design

Compare radial distribution functions

$$\theta^{(k+1)} = \theta^{(k)} - \tilde{\alpha}\int_0^{\infty} dr \left[\left(g(r\,|\,\theta) - g_{\mathrm{tgt}}(r)\right)\partial_\theta u(r\,|\,\theta)\right]_{\theta=\theta^{(k)}}$$





Lindquist-Jadrich-Truskett (2018)

Remarks

- Parameter estimation
  - Nothing but parametrised potential estimation
  - Boltzmann machine with a specific form of the potential function.
- Sampling
  - Generating samples from the probability distribution is not easy.
  - Sophisticated Monte Carlo methods (annealing)
- Quality of the estimation
  - Point estimate; most likely potential is estimated.
  - No uncertainty quantification
  - Stable to the change of parameters?

Bayesian modelling of partial differential equations (BM-PDE)

- forward problem
  - Stationary pattern generated by partial differential equations

initial condition $\psi_0(\mathbf{x})$ → PDE model $\partial_t \psi(\mathbf{x},t) = f\left(\psi(\mathbf{x},t)\right)$ → stationary pattern $\psi_s(\mathbf{x})$

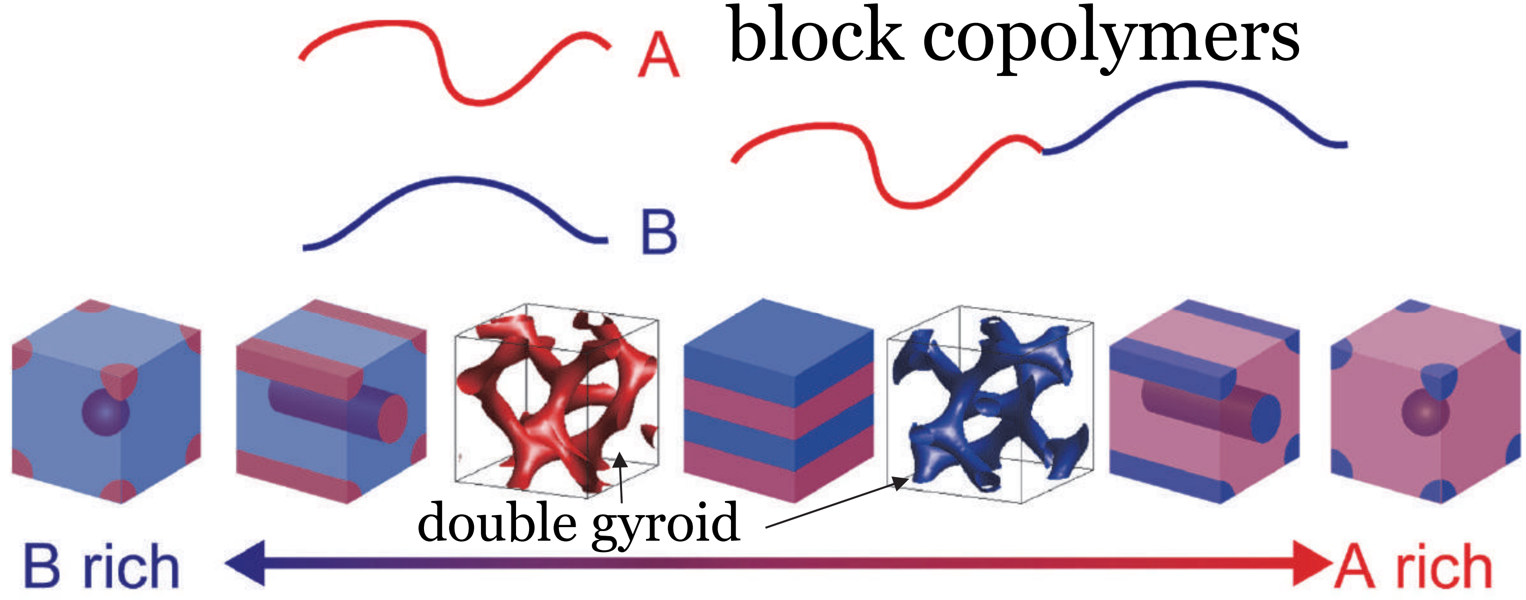
- inverse problem

target pattern $\psi^*(\mathbf{x})$ → (estimation) → PDE model $\partial_t \psi(\mathbf{x},t) = f_\mu\left(\psi(\mathbf{x},t)\right)$

$\psi^*(\mathbf{x}) \approx \psi_{s,\mu}(\mathbf{x})$ up to symmetry transformation

Pattern Formation

block copolymers



phenomenological model

partial differential equations

$$\frac{\partial \psi}{\partial t} = \Delta \frac{\delta \mathcal{F}}{\delta \psi}$$

free energy functional

$$\mathcal{F} = \int_{\Omega} \left[\frac{1}{2} |\nabla \psi|^2 + f_{\mathrm{GL}} + \frac{\alpha}{2} \int_{\Omega'} \left(\psi(\mathbf{x}) - \overline{\psi} \right) G(\mathbf{x}, \mathbf{x}') \left(\psi(\mathbf{x}') - \overline{\psi} \right) \right]$$

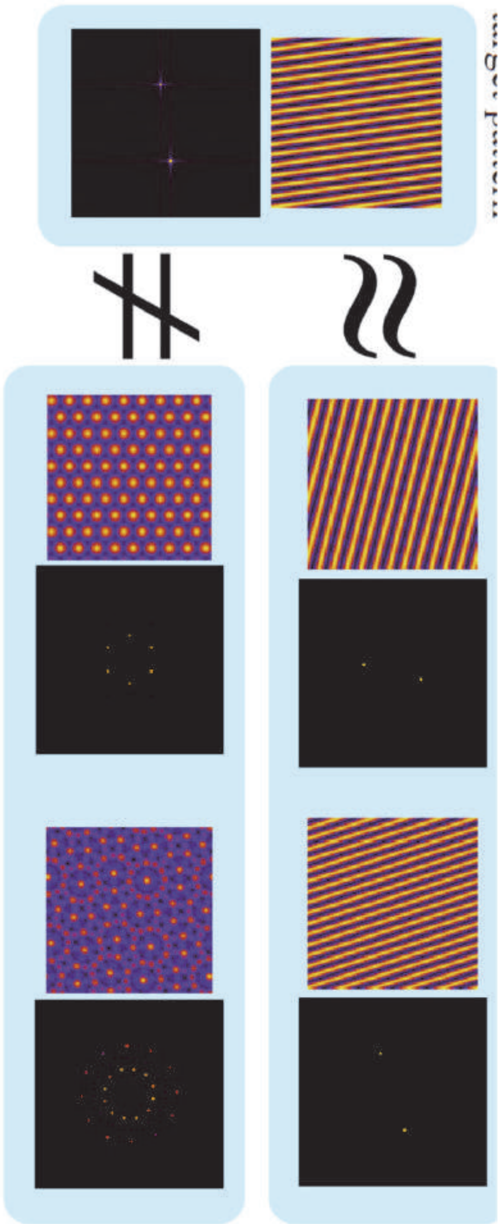
wide application

- reaction-diffusion
- fluid
- phase-field model
- crystals

(phase-field crystal)

Order Parameters

- Requirement
  - Patterns are invariant under translation and rotation.
  - Patterns respect their symmetry.
  - Small noise must be eliminated.



- Order Parameter
  - Fourier transformation $\psi_k = \mathcal{F}\left[\psi(\mathbf{x})\right]$
  - Expand the coefficients

2D $A_l = \int_{\mathbf{k}} \psi_{\mathbf{k}} e^{il\theta_k}$

2D invariant $\Psi_l = \left|A_l\right|$

$$\mathbf{k} = k\begin{pmatrix}\cos\theta_k \\ \sin\theta_k\end{pmatrix}$$

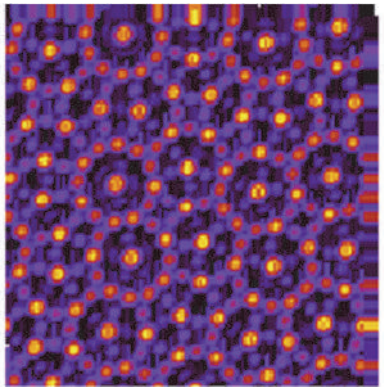
Can we estimate the best model from a snapshot?

- Very limited information
  - No information about time series
  - Symmetry is not known.
  - Not unique; Regularisation is necessary.
  - Quantify quality of the estimation; we don't know the ground-truth.

2D quasi crystal

pattern

$\psi^*(\mathbf{x})$



?

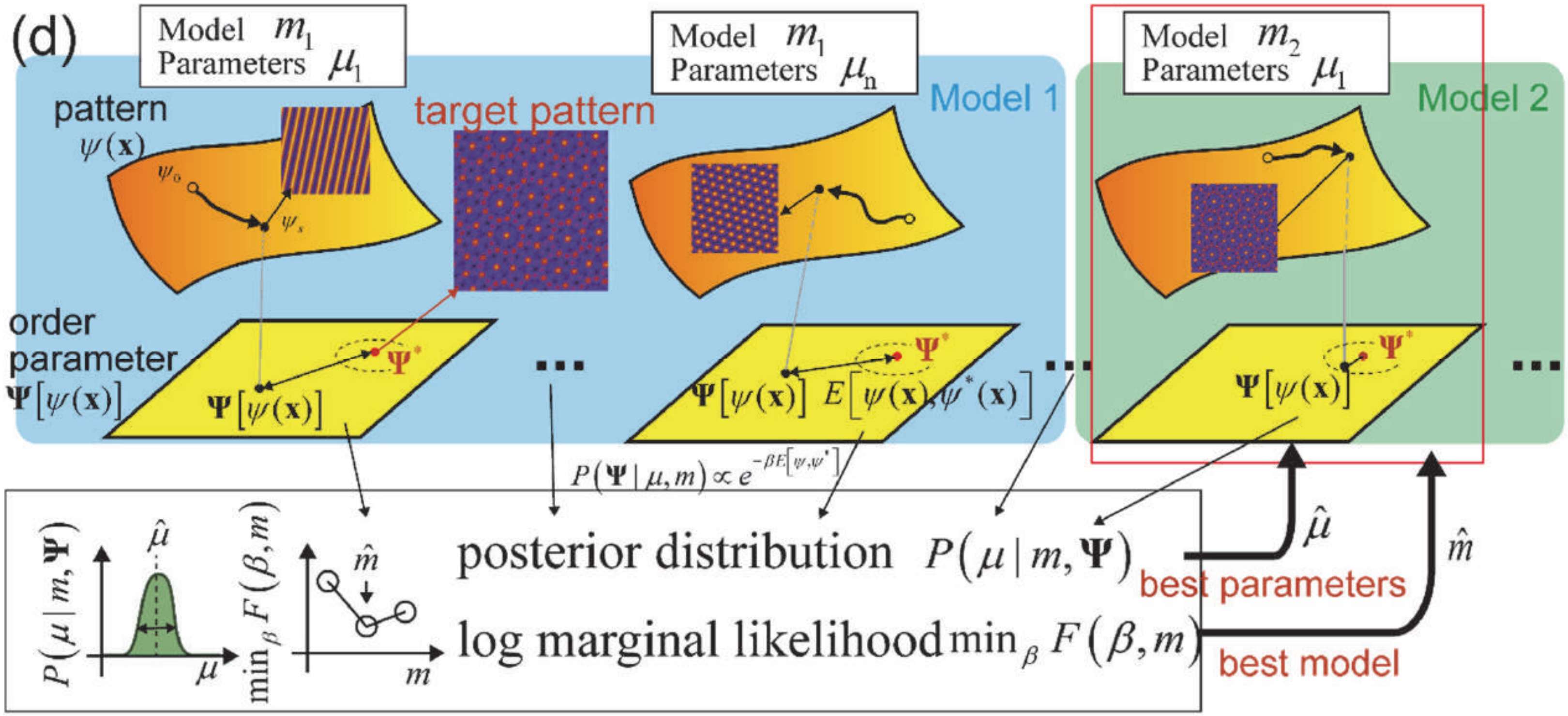
model

$$\frac{\partial \psi(\mathbf{x},t)}{\partial t} = \Delta \frac{\delta \mathcal{F}_\mu[\psi(\mathbf{x},t)]}{\delta \psi}$$

μ parameters

NY-Tokuda (2020)

Bayesian modelling of partial differential equations (BM-PDE)



- Make a family of PDE models (the number of length scales)

- Bayesian modelling + sampling by replica exchange Monte Carlo method

Parameter estimation from the posterior distribution, and model selection by marginal likelihood

Phase field crystal model

Nonlinear partial differential equations
(gradient system with mass conservation)

$$\partial_t \psi = \Delta \frac{\delta \mathcal{F}}{\delta \psi}$$

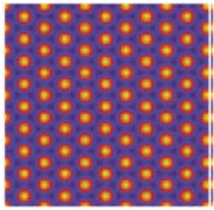
Swift-Hohenberg-type free energy

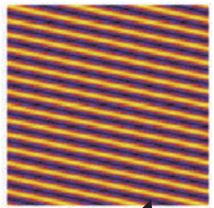
$$\mathcal{F}[\psi] = \int_\Omega \left[-\frac{\epsilon}{2}\psi^2 + \frac{1}{2}\psi(\Delta+1)^2(\Delta+q^2)^2\psi - \frac{\alpha}{3}\psi^3 + \frac{1}{4}\psi^4 \right]$$

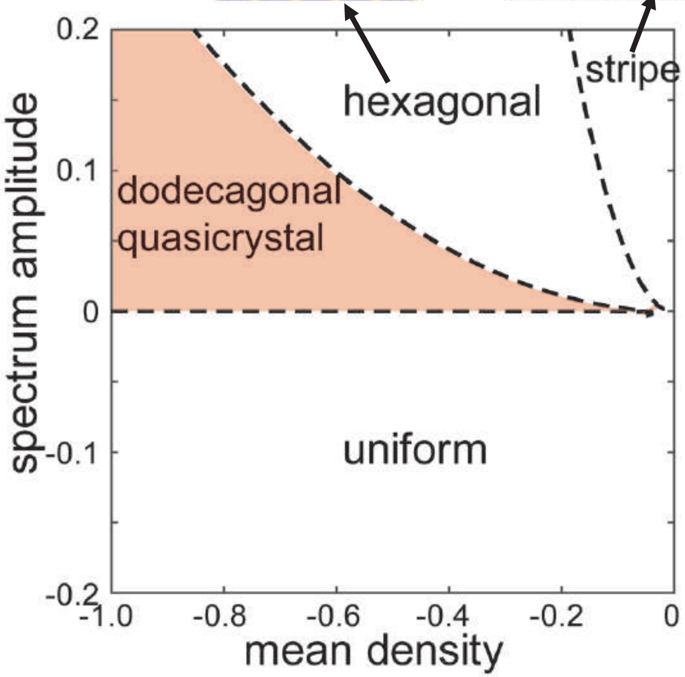
second length scale → quasicrystal

Elder, et al (2002)

For quasicrystals
Lifshitz-Petrich (1997)

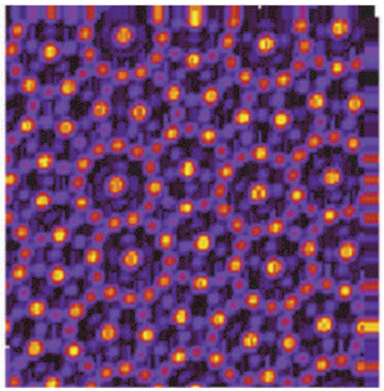






dodecagonal pattern

$q = 2\cos(\pi/12)$



$N = 128$

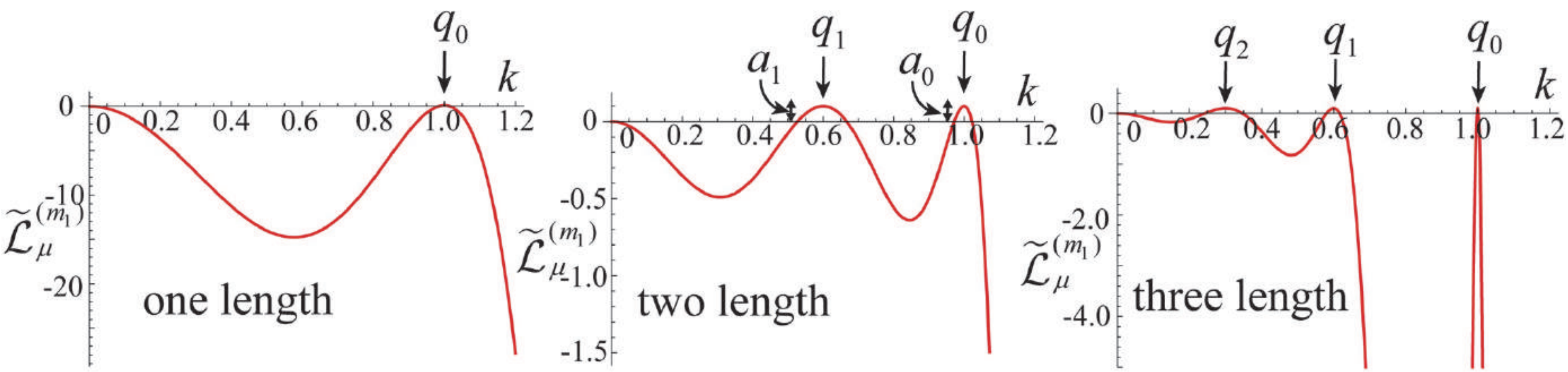
Family of models

- Family of linear operator
  - Parametrised by characteristic wave numbers and their eigenvalues

$$\partial_t \phi_k = \widetilde{\mathcal{L}}^{(m)}_{\mu,k}[\phi_k] + \mathcal{N}_k$$

$$\widetilde{\mathcal{L}}^{(m_1)}_{\mu,k}[\phi_k] = -\frac{a_0}{q_0^4} k^2 \left(k^2 - 2q_0^2\right) - \frac{s_0}{q_0^4} k^2 \left(q_0^2 - k^2\right)^2$$

one length scale two length scale three length scale



Bayes theorem

$$\text{posterior distribution} = \frac{\text{likelihood} \times \text{prior distribution}}{\text{normalisation constant}}$$

target pattern

generated pattern

parameters

parameter estimation

$$p\left(\mu \mid \Psi^*, \beta, m\right) = \frac{p\left(\Psi^* \mid \Psi_\mu, \beta, m\right) p(\mu \mid m)}{p\left(\Psi^* \mid \beta, m\right)}$$

noise level
(inverse temperature)

model selection

model

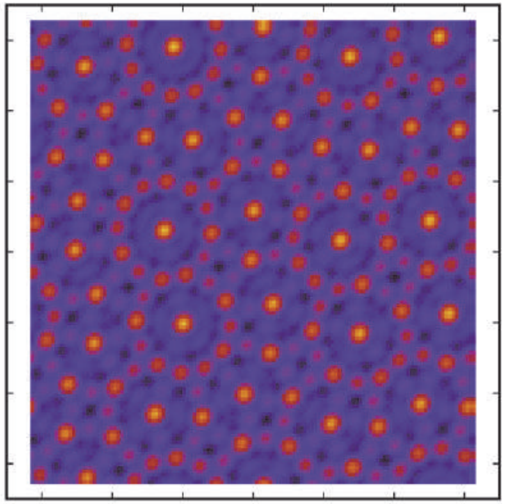
observation noise

$$\Psi^* = \Psi + \text{noise}$$

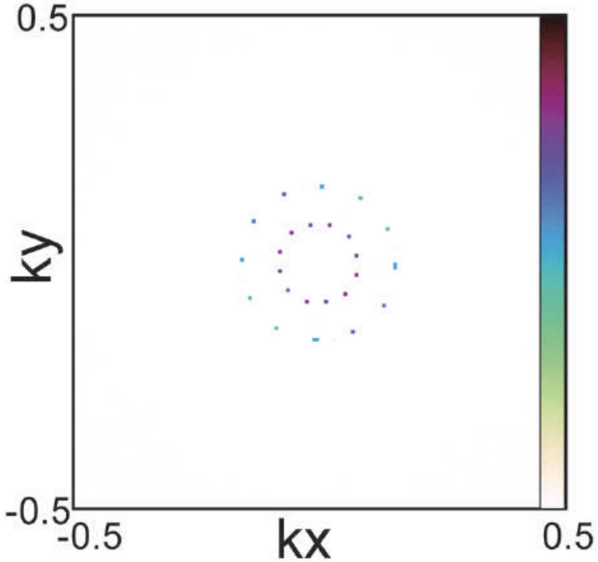
→ uncertainty quantification
(best noise level and uncertainty of parameters)

Generating a pattern through the estimation

target pattern
(real space)



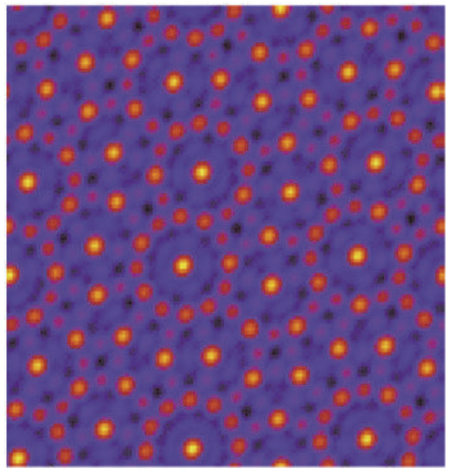
target pattern
(Fourier space)



12 spots

dodecagonal quasicrystal

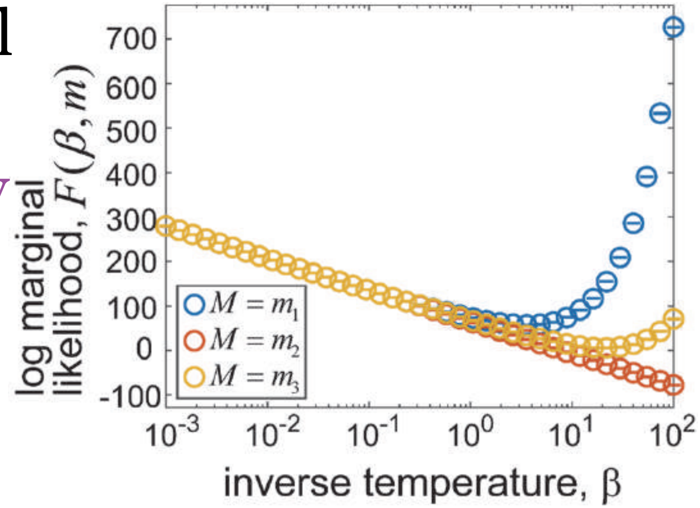
generated pattern
(estimated model
and parameters)



Model Selection and Parameter Estimation

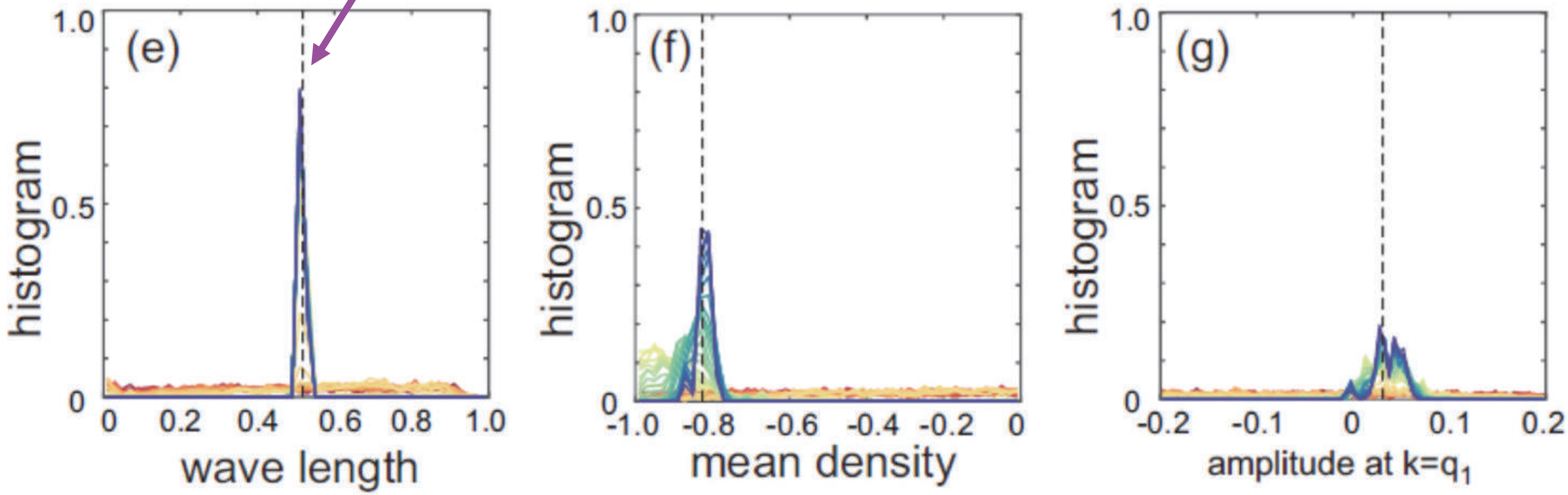
model selection from log marginal likelihood

→ two-length scales are necessary



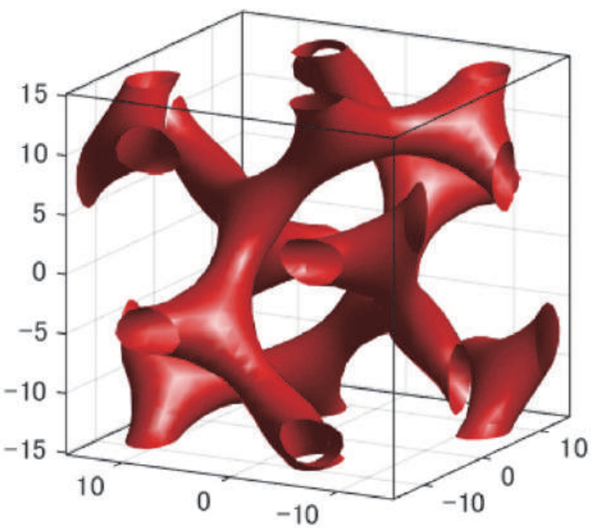
parameter estimation from posterior distribution

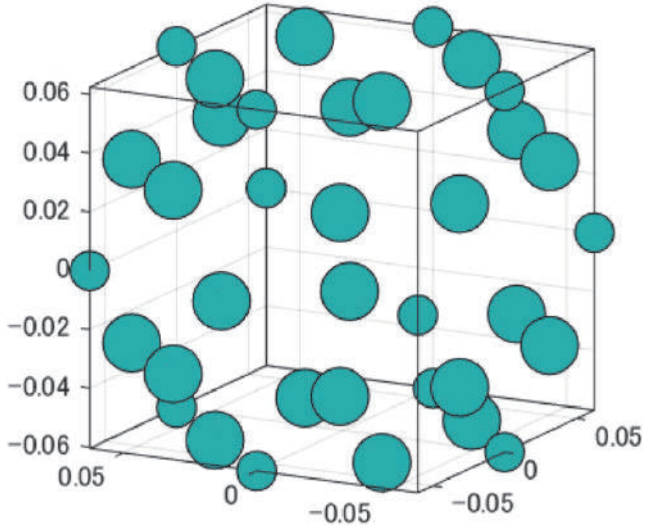
ground truth

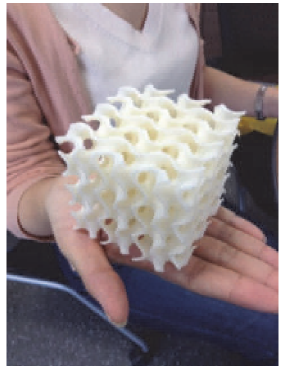


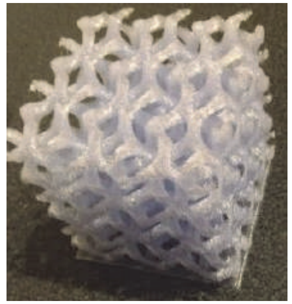
Three-Dimensional Patterns

double gyroid generated from a function

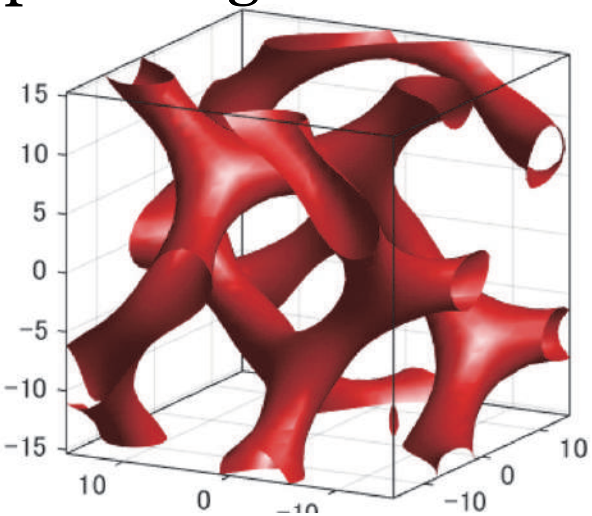


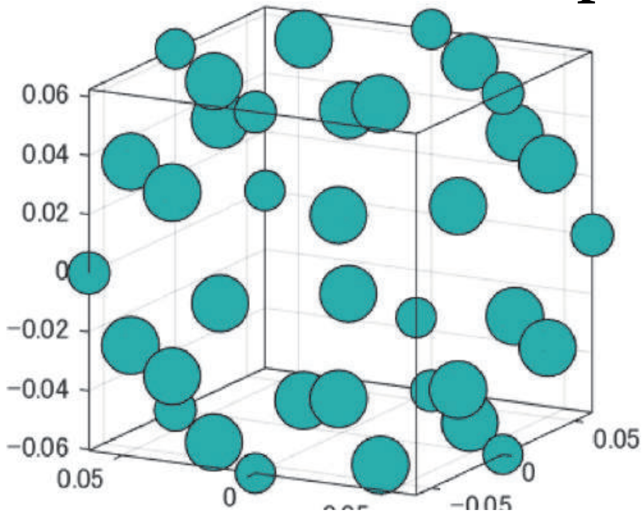




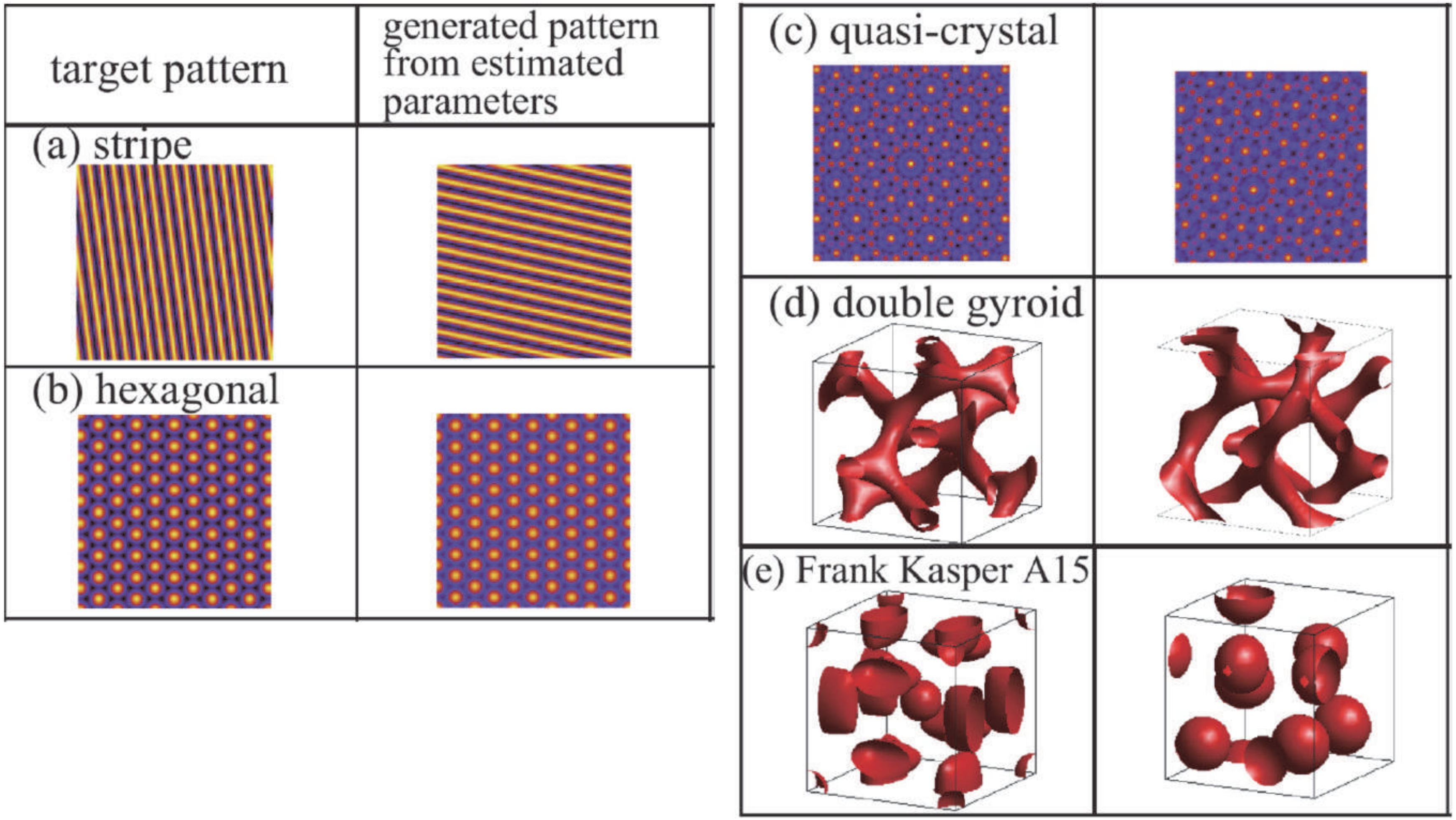


pattern generated from estimated model and parameters





Various patterns in 2D and 3D



NY-S.Tokuda (2020)

Summary

- Discovery of a governing equation
  - Not difficult than it looks
  - Becoming feasible
  - Yet, we need quantification of its quality

- Bayesian modelling of partial differential equations
  - BM-PDE works well for various patterns.
  - It works not only for a target pattern with the ground-truth, but also for a pattern without it.
  - It is also robust against noise (~20%).

Natsuhiko Yoshinaga and Satoru Tokuda
“Bayesian modelling of partial differential equations from one snapshot of pattern“, arxiv:2006.06125